

Non-Markovian property of afterpulsing effect in single-photon avalanche detector

Fang-Xiang Wang, Wei Chen,* Ya-Ping Li, De-Yong He, Chao Wang,
Yun-Guang Han, Shuang Wang, Zhen-Qiang Yin and Zheng-Fu Han

*Key Laboratory of Quantum Information,
University of Science and Technology of China,
Hefei 230026, China*

*and Synergetic Innovation Center of Quantum Information & Quantum Physics,
University of Science and Technology of China,
Hefei, Anhui 230026, China*

The single-photon avalanche photodiode (SPAD) has been widely used in research on quantum optics. The afterpulsing effect, which is an intrinsic character of SPAD, affects the system performance in most experiments and needs to be carefully handled. For a long time, afterpulsing has been presumed to be determined by the pre-ignition avalanche. We studied the afterpulsing effect of a commercial InGaAs/InP SPAD (The avalanche photodiode model is: Princeton Light-wave PGA-300) and demonstrated that its afterpulsing is non-Markovian, with a memory effect in the avalanching history. Theoretical analysis and experimental results clearly indicate that the embodiment of this memory effect is the afterpulsing probability, which increases as the number of ignition-avalanche pulses increase. This conclusion makes the principle of the afterpulsing effect clearer and is instructive to the manufacturing processes and afterpulsing evaluation of high-count-rate SPADs. It can also be regarded as a fundamental premise to handle the afterpulsing signals in many applications, such as quantum communication and quantum random number generation.

PACS numbers:

I. INTRODUCTION

Avalanche photodiodes (APDs) are widely used in a variety of fields, such as single molecule detection [1, 2], autocorrelated fluorescent decays and quantum information [3, 4], to detect extremely weak light signals or even single photons. The latter is usually named a single-photon avalanche detector (SPAD), which is an essential device and may greatly influence the performance of the system in many applications, especially in quantum information fields [4, 5]. SPADs usually work in two different modes: free-running mode and gated mode. For free-running mode, the SPAD is biased above its breakdown voltage to allow an avalanche to occur. For gated mode, the SPAD is commonly biased below its breakdown voltage, and the avalanche process can only happen in gating time windows when additional gate voltage signals are superimposed. There are many parameters that can be used to quantify the performance of SPADs, such as the clocking rate, dark count rate, detection efficiency and timing jitter. In the past decade, the performance of SPADs has been greatly improved and they can work at a gigahertz clocking frequency with up to one hundred megahertz counting rates [6–8].

The afterpulsing effect of SPADs severely limits the count rate and hinders applications that require precise measurements. Afterpulsing, which is correlated with the ignition avalanche, comes from the detrapping of carriers

that were trapped at a deep energy level in the junction depletion region [9–12]. The trapped carriers are then released from the traps over time by thermal fluctuation. Usually, the release lifetime τ is dependent on the type of trap and can vary from 10 ns to several microseconds [10, 13, 14], which is usually much longer than that needed for quenching an avalanche. A releasing carrier may initiate another pseudo detection signal, which is usually called an afterpulse, if the SPAD is biased above the breakdown voltage. The pseudo signal leads to negative effects on SPAD's applications. For example, afterpulsing will increase the quantum bit error rate in the quantum key distribution [15, 16] and will destroy the randomness of a quantum random number generator [17]. To reduce afterpulsing, another hold-off time depending on τ is necessary. The afterpulsing effect is negligible if the hold-off time is much longer than the carrier lifetime τ , but this is not practical for high-count-rate SPADs. Many approaches have been proposed to reduce afterpulsing [4, 6, 11, 18–20], but it remains a critical limitation for high-count-rate SPADs. Precisely modeling the behavior of afterpulsing is the necessary foundation to address this problem.

In this paper, we studied the afterpulsing influence of multi-ignition avalanches on the photon detection process of InGaAs/InP SPADs under different gate frequencies and photon intensities and found that afterpulsing possesses a memory effect and thus correlates to the avalanching history. We call this property as non-Markovian property [21, 22]. We developed analytical expressions for the afterpulsing influence on photon detection, enabling it to be quantified. We also conducted ex-

*weich@ustc.edu.cn

periments to verify the theoretical relationships and the experimental results fit well with the theoretical model. The model is instructive to manufacturing processes and the afterpulsing evaluation of high-count-rate SPADs. It also offers a new perspective to address the practical security of quantum communication with imperfect devices [23–25].

II. AFTERPULSING EFFECT ANALYSIS

A. Trapped carriers and afterpulsing

The SPAD (Fig. 1(a)) absorbs an incident photon in the depletion region (I region) and creates an electron-hole pair [3]. The electron then drifts into the avalanche region (P region) and triggers an avalanche. The avalanche amplifies the photocurrent so that the initial absorption event can be detected. Some carriers (electrons or holes) are trapped during the avalanche process due to material defects. The release of a carrier in the P region may initiate an afterpulse when the SPAD is biased above the breakdown voltage.

The trap lifetime τ of carriers in a SPAD can be described by the Arrhenius equation

$$\tau = C e^{E_a/kT} \quad (1)$$

where E_a is the activation energy of the trap, T is the temperature, k is the Boltzmann constant and C is determined by the relevant effective state density, cross-section of the trap and temperature T [14]. The carrier release probability is then exponential with respect to the time t . As the types of traps in APD are dependent on practical manufacturing processes, there are thereby a variety of traps with different lifetimes τ_j . Thus, the afterpulsing probability (AP) during a detection window Δt contains different exponential components [10]

$$p_a(j) = A_1 \exp(-\frac{j \cdot \Delta t}{\tau_1}) + A_2 \exp(-\frac{j \cdot \Delta t}{\tau_2}) + \dots \quad (2)$$

where A_i are the amplitudes of different exponential components, and $p_a(j)$ means the AP of the j -th detection window after the ignition avalanche. Recently, different fitting methods have been studied that employ a broad distribution of different exponential components instead of summing a few dominant discrete ones [26, 27].

So far, most studies on the afterpulsing effect just have focused on its time-dependent probability distribution ignited by a single light pulse and do not discriminate the situation of a single ignition avalanche from that with multiple avalanches [12, 28–31]. However, this is not the real situation when the SPAD works continuously for photon detection. In this case, several avalanches may happen successively within the lifetimes of the trapped carriers.

Previous works have demonstrated that the number of traps filled with carriers during an avalanche event is

much smaller than the total number [10, 14, 32]. Thus, avalanches happening successively will increase the number of filled traps. Fig. 1(b) gives a simplified view of the filling process. Different photon pulses will be absorbed by different atoms in the I region and the initial electron then drifts into P region. Here, the initial electron randomly collides atoms and each collision may create another unbound electron. The collision makes the paths of the initial and newly unbound electrons unpredictable. These unbound electrons continue the collision process until the edge of the P region. Some defects may be filled by electrons during transfers and collisions. As the filling is rare and the collisions are unpredictable, filling processes of different avalanches affect each other little, as does the release processes.

B. Non-Markovian property

As discussed above, it is reasonable to assume that the filling and release processes of different traps are independent. Thus, avalanches occurring closely will increase the number of trapped carriers and hence will enhance the AP. That is, the AP distribution will be correlated to the avalanching history, while prior studies have not delved down into this physical procedure. In the following paragraphs of this section, we will analyze the AP distribution of an SPAD that avalanches continuously within the lifetime of the dominant carriers.

Considering that the trap level distribution is a technological parameter, which is determined once the product is fabricated and the details have been studied well, we do not consider the specific time-dependent distribution of the AP in the next analysis. The primary data of AP distribution can be obtained by experimental measurements.

The probability of an avalanche triggering by a laser pulse and dark count can be derived from the Poisson distribution, and the avalanche probabilities of different gates (detection windows) have an independent identical distribution (IID) [17]. Let $p = 1 - e^{-\eta\lambda}$ be the photon-ignition avalanche probability per gate, $p_a(j)$ be the single-ignition AP of Gate- j and p_n be the total avalanche probability of Gate- n , where η is the detection efficiency of the SPAD and λ is the average photon number per laser pulse. "Single-ignition" means that for a sequence of detection gates, only the first gate (Gate-0) is triggered by the laser pulse. Thus, there is no afterpulsing for Gate-0 and $p_0 = p$.

The AP is determined by the avalanching history. The full-expression of p_n , which depends on the avalanching history, is very complex. For example, if Gate-0 avalanches, its afterpulsing effect will affect all the following gates. If Gate- j avalanches, its afterpulsing effect will affect the gates after it. The full AP expressions of Gate-1 (p_1) and Gate-2 (p_2) are as follows:

$$p_1 = p[p + (1-p)p_a(1)] + (1-p)p = p[1 + (1-p)p_a(1)], \quad (3)$$

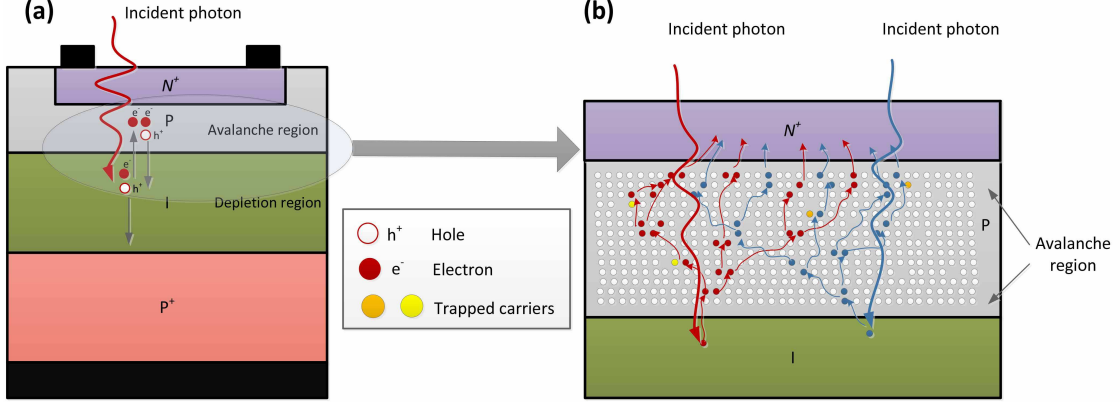


FIG. 1: (Color online) (a) Schematic drawing of a SPAD. (b) Enlarge drawing of the elliptic region in (a). It gives a simplified view of the avalanche processes and the increase of carriers (only electrons are shown here). The different colors represent different avalanche processes.

$$\begin{aligned}
 p_2 = & p[p + (1-p)p_a(1)]\{p + (1-p)[p_a(1) + p_a(2)]\} \\
 & + p\{1 - [p + (1-p)p_a(1)]\}[p + (1-p)p_a(2)] \\
 & + (1-p)p_1.
 \end{aligned} \tag{4}$$

Here we give a formal expression of p_n , and then give the approximate equation of p_n . p_n can be classified into two sets: Gate-0 avalanches in the first set $F(p, p_a(j))$ and does not avalanche in the second set $S(p, p_a(j))$. The second set can be expressed as $S(p, p_a(j)) = (1-p)p_{n-1}$. p_n can then be expressed as

$$p_n = F(p, p_a(j)) + S(p, p_a(j)) \tag{5}$$

If the single-ignition AP $p_a(j)$ is small, and only the first order terms are reserved, namely, $p_a(i)p_a(j) \cong 0$, then we obtain the approximate equation of Equation 5,

$$p_n \cong p[1 + \sum_{j=1}^n (1-p)p_a(j)]. \tag{6}$$

See Appendix for a detailed proof of Equation 6. In fact, the formal expression is enough for the proof.

If the AP is only determined by the latest ignition avalanche [33], the afterpulsing is Markovian (has no memory of prior avalanching history). That is, if Gate- $(n-1)$ avalanches, $p(n|n-1) = p + (1-p)p_a(1)$. If Gate- $(n-1)$ does not avalanche, Gate- $(n-2)$ is considered, and so on. Then the approximate equation of p_n becomes

$$p_n \cong p[1 + \sum_{j=1}^n (1-p)^j p_a(j)]. \tag{7}$$

See Appendix for a detailed proof of Equation 7.

Comparing Equation 6 with Equation 7, we find that the non-Markovian property makes the afterpulsing effect more significant.

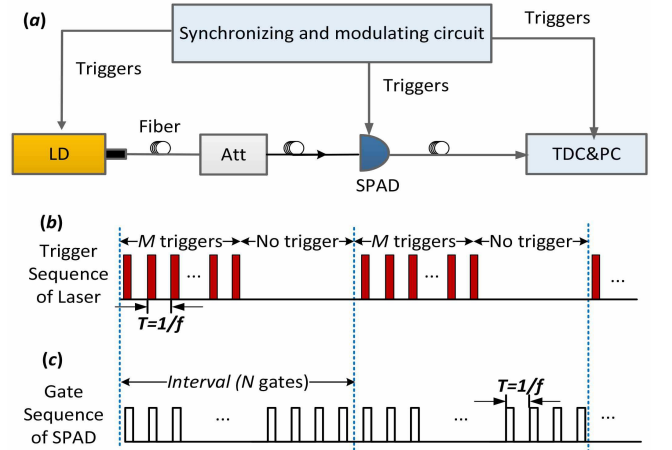


FIG. 2: (Color online) Schematic measurement setup of non-Markov property. LD: laser diode; Att: adjustable attenuator; OF: optical fiber; TDC: time-to-digital converter; PC: personal computer.

III. EXPERIMENTAL SETUP

In this section, we describe experiments to verify the non-Markovian property of afterpulsing.

The experimental setup is shown in Fig. 2 (a). The laser diode (LD) is operated in pulsed mode. Laser pulses from the LD are attenuated by an adjustable attenuator and are then sent into the SPAD. The SPAD (Princeton Lightwave PGA-300, the operating temperature is -50°C .) is operated in gated mode without dead time and the gate width is 2.5 ns . Its detection efficiency η is approximately 10.5%. The SPAD outputs a detection signal if an avalanche occurs. The detection signals are input into a time-to-digital converter (TDC, Agilent U1051A Acqiris TC890) in real time. The TDC records the arrival time of every detection signal with a resolution of 50 ps . The LD, SPAD and time-to-digital converter (TDC)

are triggered and synchronized by a synchronizing and modulating circuit.

Eq. 6 indicates that the non-Markov effect of afterpulsing will be more significant if the single-ignition AP $p_a(j)$ or light intensity increases. To verify Eq. 6, we should measure first the single-ignition AP distribution $p_a(j)$ and second the total avalanche probability p_n of Gate- n with multi-ignition pulses. Fig. 2(b) shows the modulating signal of the LD. The number of trigger pulses M (and thus the number of light pulses per interval), trigger frequency f and interval time ΔT are adjustable. $M = 1$ corresponds to the single-ignition case. To measure the AP distribution, the time with no trigger of every interval should be much longer than the carriers' lifetime τ , that is, $t_{\text{interval}} - M/f \gg \tau$. The SPAD is triggered continuously by an external pulsed signal (Fig. 2(c)) with frequency f . Parameter p can be modulated by adjusting the attenuator, and parameter $p_a(j)$ can be modulated by changing the trigger frequency f while keeping the light intensity per pulse constant.

The experiment consists of three scenarios. In **Scenario (a)**, single-ignition ($M = 1$) AP distributions under different trigger frequencies ($f = 2, 5, 10$ and 20 MHz) are measured. The average photon number λ per pulse is set as 0.07 . The AP $p_a(j)$ is the avalanche probability of Gate- j after subtracting the dark count probability.

In **Scenario (b)**, the avalanche probabilities of Gate- n p_n with different average photon numbers per pulse λ ($\lambda = 1.0, 0.7, 0.07$ and 0.02) are measured, while the AP distribution $p_a(j)$ varies little. This scenario is used to see how p_n changes with p ($p = 1 - e^{-\eta\lambda}$). As a contrast, the slope of p_n will decay faster via p if the afterpulsing is Markovian.

In **Scenario (c)**, the AP distribution $p_a(j)$ is modulated, while λ is kept constant. As discussed in Section II, the original AP distribution is given by the device technology and operation conditions and does not change. However, according to Equation 2, the value of AP with time, $p_a(j)$, can be changed by re-sampling the detection period Δt . Hence, in this scenario, the trigger frequency f is modulated ($f = 2, 5, 10$ and 20 MHz) so that the value of $p_a(j)$ becomes different, while λ is set as constant ($\lambda = 0.07$). This scenario is used to see how p_n changes with $p_a(j)$. This scenario is also used to measure the multi-ignition AP distribution $p_a(j)$, where $p_a(j)$ is the avalanche probability of Gate- $(M + j - 1)$. For all scenarios, the dark count rates per gate are under the order of 10^{-5} .

IV. RESULTS AND DISCUSSION

Single and multi-ignition AP distributions with different triggering frequency f are shown in Fig. 3. The experimental data demonstrate that AP increases as f increases, where $\lambda = 0.07$ for all cases. As shown in Fig. 3(a), the total single-ignition AP $p_a = \sum_j p_a(j) \approx 2.0\%$ when $f = 2$ MHz, while $p_a \approx 29.9\%$ when $f = 20$ MHz.

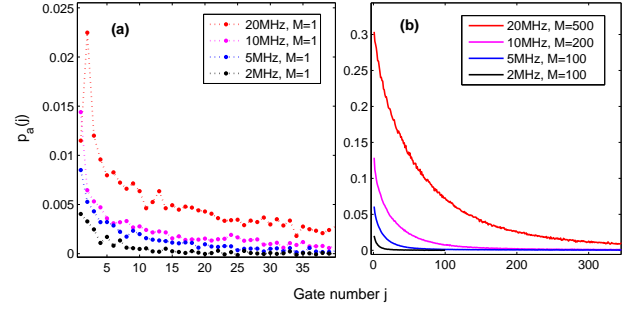


FIG. 3: (Color online) Afterpulsing probability distributions with (a) single and (b) multi-ignition pulses. The average photon number per laser pulse is $\lambda = 0.07$.

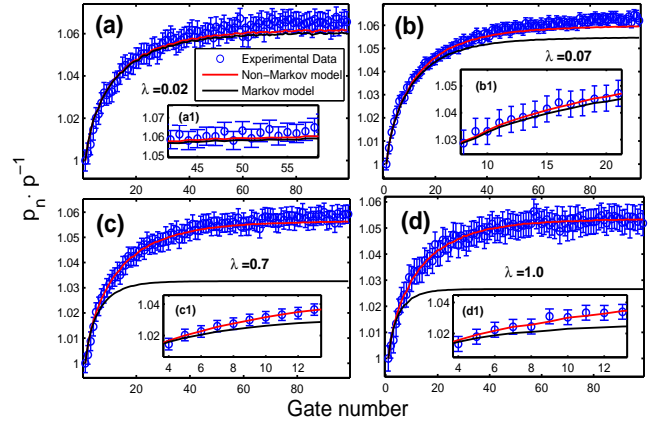


FIG. 4: (Color online) Experimental data and theoretical simulations for p_n with different average photon numbers. The trigger frequency is $f = 5$ MHz. The insets are partial enlargements. The error bar is set as $\pm 3\sigma$, where σ is the standard deviation of the statistics.

It also demonstrates that multi-ignition APs (Fig. 3(b)) are much larger than the single-ignition ones (Fig. 3(a)). For example, when $f = 5$ MHz, the multi-ignition AP of Gate- $(M + j - 1)$ is $p_{a,100}(1) = 6.1\%$, while the corresponding single-ignition AP is $p_a(1) = 0.9\%$. In fact, the total single-ignition AP $p_a \approx \sum_{j=1}^{M=100} p_a(j) = 6.2\% \approx p_{a,100}(1)$. This means that the AP accumulates as M increases. Namely, AP depends on the avalanching history and is non-Markovian. The AP becomes larger as f increases. As the full width at half maximum (FWHM) of the detection signal output from the SPAD is approximately 40 ns, the saturated processing rate of the discrimination circuit is less than 20 MHz by taking the processing time into account. Thus, $p_a(1)$ is measured to be smaller than the actual value when $f = 20$ MHz. However, the processing rate limitation is not that significant for the experiments because the average photon number $\lambda = 0.07$, which is small, when $f = 20$ MHz.

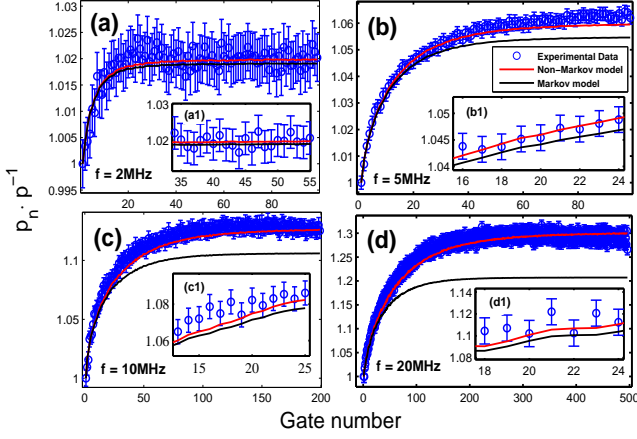


FIG. 5: (Color online) Experimental data and theoretical simulations for p_n with different gate frequency f . The average photon number per laser pulse is $\lambda = 0.07$. The detection efficiencies are $\eta_{f=2\text{MHz}} = 10.3\%$, $\eta_{f=5\text{MHz}} = 10.5\%$, $\eta_{f=10\text{MHz}} = 11.6\%$, $\eta_{f=20\text{MHz}} = 12.0\%$. The insets are partial enlargements. The error bar is also set as $\pm 3\sigma$.

Scenario (b) (see Fig. 4) shows that experimental data (the blue circles) fit Eq. 6 (the red lines) well. It proves that the approximate equation is sufficient to describe the non-Markovian property of afterpulsing. As predicted by Eq. 7, the slope of p_n (the black lines) of the Markovian model decays faster than that of the experimental data. The difference between the non-Markovian model and Markovian model is negligible when the photon-ignition avalanche probability p ($\lambda = 0.02$, $p = 1 - e^{-\eta\lambda} \approx 0.002$) is small enough. The difference becomes more significant when p increases, e. g., Fig. 4(d) shows that $p_{n=99}^{\text{non-Markov}} - p_{n=99}^{\text{Markov}} = 1.053 \cdot p - 1.027 \cdot p = 0.026 \times p$. The Markovian model cannot fit with the experimental data.

Scenario (c) provides differences of p_n between the non-Markovian and Markovian models with different $p_a(j)$ (see Fig. 5). There is hardly any difference between these two model when $f = 2\text{ MHz}$ ($p_a = 2.0\%$), as shown in Fig. 5(a). However, the difference becomes $p_{n=99}^{\text{non-Markov}} - p_{n=99}^{\text{Markov}} \approx 0.1 \cdot p$ for large n when $f = 20\text{ MHz}$ ($p_a = 29.9\%$).

The experimental results above are all well consistent with the non-Markovian analysis. According to Eq. 6, the afterpulsing influence on photon detection depends on the product of the photon-ignition avalanche probability and the total AP. In both Fig. 4(a) and Fig. 5(a), $p \cdot p_a$ is very small (approximately 10^{-4}), and the Markov model also fits well with the experimental data, although the assumption is incorrect. However, in Fig. 4(b),(c),(d) and 5(b),(c),(d), the Markovian model can no longer fit with the data, as $p \cdot p_a$ becomes larger (about 10^{-3} in 4(d) and 5(d)). The non-Markovian effect of afterpulsing will be more significant for higher p_a . This is really the case for high-count-rate SPADs.

It should be noted that, due to the limit of the de-

signed processing circuit, the detection efficiency of the SPAD varies with the trigger frequency f . To make a comparison of the AP influence with different frequencies, the gate voltage loaded on the SPAD is fine adjusted so that detection efficiency varies less ($\eta_{f=2\text{MHz}} = 10.3\%$, $\eta_{f=5\text{MHz}} = 10.5\%$, $\eta_{f=10\text{MHz}} = 11.6\%$, $\eta_{f=20\text{MHz}} = 12.0\%$). Otherwise, afterpulsing would be a catastrophe and the SPAD under $f = 20\text{ MHz}$ could no longer work properly if the voltage remains the same as that under $f = 2\text{ MHz}$. By utilizing high-speed (e.g., 100 MHz) processing circuit with improved design, the detection efficiency will vary less under the same conditions.

Early works have demonstrated that the number of traps filled with carriers during an avalanche event is much smaller than the total number in both visible and near-infrared APD [10, 14, 32]. Thus, the afterpulsing effects of these types of SPADs can be described by the non-Markovian model shown in Figure 1. Recently, afterpulsing effect is still a serious limit toward high-count-rate SPADs [18, 19]. The SPADs with gigahertz clocking frequency and one hundred megahertz counting rate are mainly reached by circuit improvement, by which the gate voltage, total avalanche current and the AP have been suppressed significantly [6, 7, 20]. However, the material quality has not been improved much in the last decade [34] and the defects of the materials are still in large quantity [18]. Hence, despite that only a particular commercial InGaAs/InP SPAD has been verified in this paper, the non-Markovian model is general for SPADs unless a significant breakthrough of the fabrication technology is being made. Our work reveals that the circuit quenching method, which cannot change the intrinsic non-Markovian property of afterpulsing, will remain one of the bottlenecks for higher-count-rate SPADs. The once and for all solution is to reduce the defects in the materials or to erase the trapped carriers before the arrival of the next photon.

V. CONCLUSIONS

In conclusion, we have proved that the afterpulsing effect of the SPAD has a memory of the avalanching history and is hence non-Markovian. The approximate expression (Eq. 6) well describes the non-Markovian property of afterpulsing. The non-Markovian effect will be more significant for higher p_a . This is the case for high-count-rate SPADs, where the detection period is much smaller than the lifetime of the carriers [6–8]. As the embodiment of rarely filling of traps, the non-Markovian property of afterpulsing should be general to different types of SPADs provided that the filling is rare during an avalanching process. Our work makes the principle of the afterpulsing effect clearer and is instructive to manufacturing processes and the afterpulsing evaluation of high-count-rate SPADs. It also offers a new perspective to consider the practical security of quantum communication due to the imperfection of devices.

Appendix

A. The Markovian approximate equation

If AP distribution is only determined by the latest ignition avalanche as assumed in [33], namely, the AP distribution has no memory of prior avalanching history and is Markovian. That is, if Gate- $(n-1)$ avalanches, $p(n|n-1) = p + (1-p)p_a(1)$. If Gate- $(n-1)$ does not avalanche, Gate- $(n-2)$ is considered. If Gate- $(n-2)$ avalanches, $p(n|(n-1, n-2)) = p + (1-p)p_a(2)$. If Gate- $(n-2)$ does not avalanche, Gate- $(n-3)$ is considered, and so on. Hence, the avalanche probability of Gate- n is

$$p_n = p_{n-1}[p + (1-p)p_a(1)] + (1-p_{n-1})\{p_{n-2}[p + (1-p)p_a(2)] + (1-p_{n-2})\{\cdots\{p_0[p + (1-p)p_a(n)] + (1-p_0)p\}\cdots\}\}. \quad (1)$$

The first term of the equation means if the pre-gate avalanches, the ones before the pre-one are not considered. If the pre-gate does not avalanche, the gate before the pre-gate is under considered, and so on.

By keeping the first-order terms of $p_a(j)$ only, we obtain the approximate equation

$$p_n \cong p[1 + \sum_{j=1}^n (1-p)^j p_a(j)]. \quad (2)$$

Proof. First, we consider the limiting case of no after-pulsing. In this case, $p_n = p$ always holds.

Thus, in general, p_n can be expressed as $p_n = p + g_n(p_a(j))$, where $g_n(p_a(j)) = g_n(p_a(1), p_a(2), p_a(3), \dots)$ is a power series of $p_a(j)$ and $j = 1, 2, \dots$. For the first term of right side, we have

$$\begin{aligned} & p_{n-1}[p + (1-p)p_a(1)] \\ &= [p + g_{n-1}(p_a(j))] \cdot [p + (1-p)p_a(1)] \\ &= p^2 + p(1-p)p_a(1) + pg_{n-1}(p_a(j)) \\ &\quad + g_{n-1}(p_a(j)) \cdot (1-p)p_a(1) \\ &\cong p^2 + p(1-p)p_a(1) + pg_{n-1}(p_a(j)). \end{aligned} \quad (3)$$

The remainder terms become

$$\begin{aligned} & (1-p_{n-1})\{p_{n-2}[p + (1-p)p_a(2)] + (1-p_{n-2}) \\ & \quad \{\cdots\{p_0[p + (1-p)p_a(n)] + (1-p_0)p\}\cdots\}\} \\ &= [1-p-g_{n-1}(p_a(j))]\{p_{n-2}[p + (1-p)p_a(2)] + (1-p_{n-2}) \\ & \quad \{\cdots\{p_0[p + (1-p)p_a(n)] + (1-p_0)p\}\cdots\}\} \\ &= (1-p)\{p_{n-2}[p + (1-p)p_a(2)] + (1-p_{n-2}) \\ & \quad \{\cdots\{p_0[p + (1-p)p_a(n)] + (1-p_0)p\}\cdots\}\} \\ & \quad - g_{n-1}(p_a(j))\{p_{n-2}[p + (1-p)p_a(2)] + (1-p_{n-2}) \\ & \quad \{\cdots\{p_0[p + (1-p)p_a(n)] + (1-p_0)p\}\cdots\}\}. \end{aligned} \quad (4)$$

In order to simplify term $g_{n-1}(p_a(j))\{p_{n-2}[p + (1-p)p_a(2)] + (1-p_{n-2})\{\cdots\{p_0[p + (1-p)p_a(n)] + (1-p_0)p\}\cdots\}\}$, we begin with the innermost brace. The

term containing $p_a(n)$ can be abandoned directly by the approximation convention $p_a(i)p_a(j) \cong 0$. Then, the expression of the innermost brace is simplified to p . Then, the expression outside the innermost brace becomes $p_1[p + (1-p)p_a(n-1)] + (1-p_1)p \cong p$. The expressions of outer braces can be simplified by analogy, and $g_{n-1}(p_a(j))\{p_{n-2}[p + (1-p)p_a(2)] + (1-p_{n-2})\{\cdots\{p_0[p + (1-p)p_a(n)] + (1-p_0)p\}\cdots\}\} \cong g_{n-1}(p_a(j))p$. Equation

1 hence becomes

$$p_n = p^2 + p(1-p)p_a(1) + (1-p)\{p_{n-2}[p + (1-p)p_a(2)] + (1-p_{n-2})\{\cdots\{p_0[p + (1-p)p_a(n)] + (1-p_0)p\}\cdots\}\} \quad (5)$$

Analogously, p_{n-2}, p_{n-3}, \dots can be simplified by the same procedure done for p_{n-1} . Equation 1 finally becomes

$$\begin{aligned} p_n &= p^2 + (1-p)p^2 + (1-p)^2p^2 + \cdots + (1-p)^{n-2}p^2 \\ &\quad + (1-p)^{n-1}p + p(1-p)p_a(1) + p(1-p)^2p_a(2) \\ &\quad + \cdots + p(1-p)^{n-1}p_a(n-1) + p_0(1-p)^n p_a(n) \\ &= p[1 + \sum_{j=1}^n (1-p)^j p_a(j)]. \end{aligned} \quad (6)$$

This completes the proof. \square

There is an alternative proof of the Markovian approximate equation. The proof is shown below.

Proof. The recurrence relation of the sequence p_n can be written as follows according to Equation 1:

$$p_{n+1} = p + p_n(1-p)p_a(1) + p_{n-1}(1-p_n)(1-p)p_a(2) + \dots + p(1-p_1)\dots(1-p_n)(1-p)p_a(n+1) \quad (7)$$

According to Equation 7, it is easy to get that $p_1 = p + p(1-p)p_a(1)$, $p_2 = p + p(1-p)p_a(1) + p(1-p_1)(1-p)p_a(2) \cong p[1 + (1-p)p_a(1) + (1-p)^2p_a(2)]$, where we only keep the first-order terms of $p_a(j)$. Thus, we assume that $p_n \cong p[1 + \sum_{j=1}^n (1-p)^j p_a(j)]$. If $p_{n+1} = p_n + p(1-p)^{n+1}p_a(n+1)$ holds, then Equation 2 holds.

According to the expression of p_n , we can replace all the p_n in Equation 7 by p since we neglect higher than the first-order terms of $p_a(j)$. Thus we can get straightforwardly:

$$\begin{aligned} p_{n+1} &\cong p + p(1-p)p_a(1) + p(1-p)^2p_a(2) \\ &\quad + \dots + p(1-p)^{n+1}p_a(n+1) \\ &= p[1 + \sum_{j=1}^{n+1} (1-p)^j p_a(j)] \end{aligned} \quad (8)$$

This means Equation 2 holds for $n+1$, thus completes the proof. \square

B. The non-Markovian approximate equation

The formal expression of avalanche probability of Gate- n in non-Markovian model is

$$p_n = F(p, p_a(j)) + S(p, p_a(j)) \quad (9)$$

By keeping the first-order terms of $p_a(j)$ only, we obtain the approximate equation of non-Markovian model

$$p_n \cong p[1 + \sum_{j=1}^n (1-p)(p_a(j))]. \quad (10)$$

Proof. According to Equation 9, it is easy to get that $p_1 = p[1 + (1-p)p_a(1)]$, $p_2 = p[1 + (1-p)(p_a(1) + p_a(2)) + (1-p)^2p_a(1)^2] \cong p[1 + (1-p)(p_a(1) + p_a(2))]$, and $p_3 = p[1 + (1-p)(p_a(1) + p_a(2) + p_a(3)) + (1-p)^2p_a(1)^2 + 2(1-p)^2p_a(1)p_a(2) + (1-p)^3p_a(1)^3] \cong p[1 + (1-p)(p_a(1) + p_a(2) + p_a(3))]$. Thus, we assume that $p_{n-1} = p[1 + (1-p)(p_a(1) + p_a(2) + \dots + p_a(n-1))]$. If $p_n = p_{n-1} + p(1-p)p_a(n)$ holds, then Equation 10 holds.

The right side of Equation 9 includes two sets: Gate-0 avalanches in the first set $F(p, p_a(j))$ and does not avalanche in the second set $S(p, p_a(j))$. The second set can be expressed as $S(p, p_a(j)) = (1-p)p_{n-1}$. Equation 9 can then be expressed as

$$\begin{aligned} p_n &= \underbrace{p}_{\text{Gate-0}} \cdot \underbrace{[p + (1-p)p_a(1)]}_{\text{Gate-1}} \cdot \underbrace{\{\dots\}}_{f_1(p, p_a(j))} \\ &\quad + \underbrace{p}_{\text{Gate-0}} \cdot \underbrace{[1 - p - (1-p)p_a(1)]}_{\text{Gate-1}} \cdot \underbrace{\{\dots\}}_{f'_1(p, p_a(j))} \quad (11) \\ &\quad + \underbrace{(1-p)}_{\text{Gate-0}} \cdot \underbrace{p_{n-1}}_{\text{Remaining terms}} \\ &\quad \underbrace{\hspace{10em}}_{S(p, p_a(j))} \end{aligned}$$

The main difference of $F(p, p_a(j))$ from $S(p, p_a(j))$ is the afterpulsing effect ignited by Gate-0. If we separate out all terms due to the afterpulsing effect of Gate-0 (that is, terms with factors $p_a(1)$ of Gate-1, $p_a(2)$ of Gate-2, \dots , $p_a(n)$ of Gate- n), the remaining terms of $F(p, p_a(j))$ can be expressed as $p \cdot p_{n-1}$. According to Equation 11, terms with factor $p_a(1)$ of Gate-1 can be expressed in the form of $p(1-p)p_a(1)f_1(p, p_a(j)) - p(1-p)p_a(1)f'_1(p, p_a(j))$, where $f_1(p, p_a(j))$ and $f'_1(p, p_a(j))$ are the abbreviations for the remaining terms. According to the approximate convention, all terms with $p_a(j)$ in $f_1(p, p_a(j))$ and $f'_1(p, p_a(j))$ are abandoned. Then $f_1(p, p_a(j))$ and $f'_1(p, p_a(j))$ approximate to the same expression

$$\begin{aligned} f_1(p, p_a(j)) &\cong f'_1(p, p_a(j)) \\ &\cong \binom{n-1}{0} p^{n-1} + \binom{n-1}{1} p^{n-2}(1-p) \\ &\quad + \dots + \binom{n-1}{n-1} (1-p)^{n-1} \\ &= [p + (1-p)]^{n-1} \\ &= 1, \end{aligned} \quad (12)$$

where $\binom{n-1}{k}$ is the binomial coefficient. Hence, $p(1-p)p_a(1)f_1(p, p_a(j)) - p(1-p)p_a(1)f'_1(p, p_a(j)) \cong 0$. Analogously, Terms with factor $p_a(2)$ of Gate-2, $p_a(3)$ of Gate-3, \dots , $p_a(n-1)$ of Gate- $(n-1)$ all approximate to 0. As there is no term $-(1-p)p_a(n)$ for Gate- n , there remains a nonzero term $p(1-p)p_a(n)f_n(p, p_a(j)) \cong p(1-p)p_a(n)$, where $f_n(p, p_a(j)) \cong 1$ by the approximation convention. Thus,

$$\begin{aligned} p_n &= F(p, p_a(j)) + S(p, p_a(j)) \\ &= p \cdot p_{n-1} + p(1-p)p_a(n) + (1-p)p_{n-1} \\ &= p_{n-1} + p(1-p)p_a(n) \quad (13) \\ &= p[1 + \sum_{j=1}^n (1-p)(p_a(j))]. \end{aligned}$$

This completes the proof. \square

-
- [1] L.-Q. Li and L. M. Davis, "Single photon avalanche diode for single molecule detection," *Rev. Sci. Instrum.*, vol. 64, no. 1993, pp. 1524-1529, 1993.
 - [2] W. E. Moerner and D. P. Fromm, "Methods of single-molecule fluorescence spectroscopy and microscopy," *Rev. Sci. Instrum.*, vol. 74, no. 8, pp. 3597-3619, 2003.
 - [3] M. D. Eisaman, J. Fan, A. Migdall, and S. V. Polyakov, "Invited Review Article: Single-photon sources and detectors," *Rev. Sci. Instrum.*, vol. 82, no. 7, pp. 071101(1-25), 2011.
 - [4] R. H. Hadfield, "Single-photon detectors for optical quantum information applications," *Nat. Photonics*, vol. 3, no. 12, pp. 696-705, Dec. 2009.
 - [5] L. C. Comandar, B. Frhlich, M. Lucamarini, K. A. Patel, A. W. Sharpe, J. F. Dynes, Z. L. Yuan, R. V. Pentty, and A. J. Shields, "Room temperature single-photon detectors for high bit rate quantum key distribution," *Appl. Phys. Lett.*, vol. 104, no. 2, pp. 021101(1-4), Jan. 2014.
 - [6] Z. L. Yuan, B. E. Kardynal, A. W. Sharpe, and A. J. Shields, "High speed single photon detection in the near infrared," *Appl. Phys. Lett.*, vol. 91, no. 4, pp. 041114(1-3), 2007.
 - [7] K. A. Patel, J. F. Dynes, A. W. Sharpe, Z. L. Yuan, R. V. Pentty, and A. J. Shields, "Gigacount/second photon detection with InGaAs avalanche photodiodes," *Electron. Lett.*, vol. 48, no. 2, pp. 111-113, 2012.
 - [8] A. Tosi, C. Scarcella, G. Boso, and F. Acerbi, "Gate-Free InGaAs/InP Single-Photon Detector Working at Up to 100 Mcount/s," *IEEE Photonics J.*, vol. 5, no. 4, pp. 6801308(1-8), Aug. 2013.
 - [9] R. H. Haitz, "Mechanisms Contributing to the Noise Pulse Rate of Avalanche Diodes," *J. Appl. Phys.*, vol. 36, no. 10, pp. 3123-3131, 1965.
 - [10] S. Cova, A. Lacaita, and G. Ripamonti, "Trapping phenomena in avalanche photodiodes on nanosecond scale," *Electron Device Lett. IEEE*, vol. 12, no. 12, pp. 685-687, 1991.
 - [11] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche photodiodes and quenching circuits for single-photon detection," *Appl. Opt.*, vol. 35, no. 12, pp. 1956-1976, Apr. 1996.
 - [12] Y. Kang, H. X. Lu, Y.-H. Lo, D. S. Bethune, and W. P. Risk, "Dark count probability and quantum efficiency of avalanche photodiodes for single-photon detection," *Appl. Phys. Lett.*, vol. 83, no. 14, pp. 2955-2957, 2003.
 - [13] A. C. Giudice, M. Ghioni, S. Cova, and F. Zappa, "A process and deep level evaluation tool: afterpulsing in avalanche junctions," in *European Solid-State Device Research, 2003. ESSDERC 03. 33rd Conference, 2003*, pp. 347-350.
 - [14] K. E. Jensen, P. I. Hopman, E. K. Duerr, E. a. Dauler, J. P. Donnelly, S. H. Groves, L. J. Mahoney, K. a. McIntosh, K. M. Molvar, A. Napoleone, D. C. Oakley, S. Vergheze, C. J. Vineis, and R. D. Younger, "Afterpulsing in Geiger-mode avalanche photodiodes for 1.06 μm wavelength," *Appl. Phys. Lett.*, vol. 88, no. 13, pp. 133503(1-3), 2006.
 - [15] A. Yoshizawa, R. Kaji, and H. Tsuchida, "After-pulse-discarding in single-photon detection to reduce bit errors in quantum key distribution," *Opt. Express*, vol. 11, no. 11, pp. 1303-1309, Jun. 2003.
 - [16] N. Jain, B. Stiller, I. Khan, V. Makarov, C. Marquardt, and G. Leuchs, "Risk Analysis of Trojan-Horse Attacks on Practical Quantum Key Distribution Systems," *IEEE J. Sel. Top. Quantum Electron.*, vol. 21, no. 3, pp. 6600710(1-10), May 2015.
 - [17] F.-X. Wang, C. Wang, W. Chen, S. Wang, F.-S. Lv, D. He, Z.-Q. Yin, H.-W. Li, G.-C. Guo, and Z.-F. Han, "Robust quantum random number generator based on avalanche photodiodes," *J. Light. Technol.*, vol. 33, no. 15, pp. 3319-3326, 2015.
 - [18] A. Tosi, N. Calandri, M. Sanzaro, and F. Acerbi, "Low-Noise, Low-Jitter, High Detection Efficiency InGaAs/InP Single-Photon Avalanche Diode," *IEEE J. Sel. Top. Quantum Electron.*, vol. 20, no. 6, pp. 192-197, Nov. 2014.
 - [19] M. A. Wayne, A. Restelli, J. C. Bienfang, and P. G. Kwiat, "Afterpulse Reduction Through Prompt Quenching in Silicon Reach-Through Single-Photon Avalanche Diodes," *J. Light. Technol.*, vol. 32, no. 21, pp. 3495-3501, 2014.
 - [20] C. Scarcella, G. Boso, A. Ruggeri, and A. Tosi, "In-GaAs/InP Single-Photon Detector Gated at 1.3 GHz With 1.5% Afterpulsing," *IEEE J. Sel. Top. Quantum Electron.*, vol. 21, no. 3, pp. 3800306(1-6), 2015.
 - [21] A. T. Bharucha-Reid, *Elements of the Theory of Markov Processes and their Applications*. Courier Corporation (2012).
 - [22] A process satisfies the Markovian property if one can make predictions for the future of the process based solely on its present state, i. e., conditional on the present state of the system, its future and past are independent. Otherwise, if its future and past are correlated, the process is non-Markov. As shown in this letter, the afterpulsing probability distribution depends on the avalanching history. Thus afterpulsing effect satisfies non-Markovian property.
 - [23] D. Gottesman, H. Lo, L. Norbert, and J. Preskill, "Security of quantum key distribution with imperfect devices," *Inf. Theory, 2004. ISIT 2004. Proceedings. Int. Symp. on*, IEEE, pp. 136-136, 2004.
 - [24] R. Renner, "Security of quantum key distribution," *Int. J. Quantum Inf.*, vol. 6, no. 1, pp. 1-127, 2008.
 - [25] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dusek, N. Ltkenhaus, and M. Peev, "The security of practical quantum key distribution," *Rev. Mod. Phys.*, vol. 81, no. 3, pp. 1301-1350, 2009.
 - [26] M. A. Itzler, X. Jiang, and M. Entwistle, "Power law temporal dependence of InGaAs/InP SPAD afterpulsing," *J. Mod. Opt.*, vol. 59, no. 17, pp. 1472-1480, Oct. 2012.
 - [27] D. B. Horoshko, V. N. Chizhevsky, and S. Y. Kilin, "Full-response characterization of afterpulsing in single-photon detectors," *arXiv Prepr. arXiv 1409.6752*, pp. 1-4, Sep. 2014.
 - [28] A. Yoshizawa, R. Kaji, and H. Tsuchida, "Quantum efficiency evaluation method for gated-mode single-photon detector," *Electron. Lett.*, vol. 38, no. 23, pp. 1468-1469, 2002.
 - [29] H. T. Yen, S. D. Lin, and C. M. Tsai, "A simple method to characterize the afterpulsing effect in single photon avalanche photodiode," *J. Appl. Phys.*, vol. 104, no. 5, pp. 054504(1-4), 2008.
 - [30] T. Ferreira da Silva, G. B. Xavier, and J. P. von

- der Weid, "Real-Time Characterization of Gated-Mode Single-Photon Detectors," *IEEE J. Quantum Electron.*, vol. 47, no. 9, pp. 1251-1256, Sep. 2011.
- [31] M. Stipcevic, D. Wang, and R. Ursin, "Characterization of a commercially available large area, high detection efficiency single-photon avalanche diode," *J. Light. Technol.*, vol. 31, no. 23, pp. 3591-3596, 2013.
- [32] Y. Liu, S. R. Forrest, J. Hladky, M. J. Lange, G. H. Olsen, and D. E. Ackley, "A planar InP/InGaAs avalanche photodiode with floating guard ring and double diffused junction," *J. Light. Technol.*, vol. 10, no. 2, pp. 182-193, 1992.
- [33] R. Humer, Gerhard and Peev, Momtchil and Schaeff, Christoph and Ramelow, Sven and Stipcevic, Mario and Ursin, "A simple and robust method for estimating afterpulsing in single photon detectors," *J. Light. Technol.*, vol. 33, no. 14, pp. 3098-3107, 2015.
- [34] M. a. Itzler, X. Jiang, M. Entwistle, K. Slomkowski, A. Tosi, F. Acerbi, F. Zappa, and S. Cova, *J. Mod. Opt.* 58, 174 (2011).